UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/782,955	02/23/2004	Chih-Chung Kuo	KUOC3019/EM	7575

23364 7590 06/11/2007

BACON & THOMAS, PLLC
625 SLATERS LANE
FOURTH FLOOR
ALEXANDRIA, VA 22314

EXAMINER	
GODBOLD, DOUGLAS	
ART UNIT	**PAPER NUMBER**
2626	
MAIL DATE	**DELIVERY MODE**
06/11/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)
	10/782,955	KUO ET AL.
	Examiner	Art Unit
	Douglas C. Godbold	2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on *23 February 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-18* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-16 and 18* is/are rejected.

7)☐ Claim(s) *17* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *23 February 2004* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *20040223*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This office action is in response to application 10/782,955 filed February 23, 2004. Claims 1-18 are pending in the application and have been examined.

Priority

2. This application claims priority to Taiwan application 092125187 filed September 12, 2003. This priority date has been considered in this application.

Information Disclosure Statement

3. The Information Disclosure Statement filed February 23, 2004 has been considered in this office action.

Claim Rejections - 35 USC § 101

4. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

5. Claims 1-9 are rejected under 35 U.S.C. 101 because the claimed invention lacks patentable utility. Claim 1 attempts to claim an automatic speech segmentation and verification method. However there is no output or tangible result of the method, causing the invention to lack utility. Therefore claim 1 is rejected and 2-9 are also rejected under 35 U.S.C. 101 as they are dependent of claim 1.

Claim Rejections - 35 USC § 102

6. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

7. Claims 1-7, 8-16 and 18 are rejected under 35 U.S.C. 102(a) as being anticipated by Kuo et al (Automatic Speech segmentation and Verification for Concatenative Synthesis).

8. Consider claim 1, Kuo teaches an automatic speech segmentation and verification method (this paper presents an automatic speech segmentation method based on HHM alignment... Abstract) comprising:

a retrieving step, for retrieving a recorded speech corpus, the recorded speech corpus corresponding to a known text script, the known text script defining phonetic information with N phonetic units (inherent as the script and speech corpus must be retrieved to be examined.);

a segmenting step, for segmenting the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script (The analysis window is 20 ms with a window shift of 10 ms. The feature vector has 26 dimensions including 12 Mel-scale cepstral coefficients (MFCC), 12 delta-cepstral coefficients, I delta-energy, and 1 delta-delta-energy. The speaker-dependent HMMs are left-to-right models including 100 3-state right context-dependent

Initial models, 38 5-state Final models, and a single I-state silence model. At first, we used speaker-independent HMMs as the initial models for training the speaker-dependent HMMs, section 2.1);

a segment-confidence-measure verifying step, for verifying segment confidence measures of N cutting points of the test speech unit segments to determine if the N cutting points of the test speech unit segments are correct (Therefore, the confidence measure for syllable segmentation (CMS) is designed according to the following factors: •Degree of disagreement among different results from multiple experts as well as from HMM. •Duration statistics for different Initials, Finals, and syllable types; section 2.3.);

a phonetic-confidence-measure verifying step, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script (The purpose of syllable verification is to check the phonetic consistence between the recorded syllable segment and the syllable type according to the text script; section 3, line 1.); and

a determining step, for determining acceptance of the phonetic unit by comparing a combination of segment reliability and the phonetic confidence measures of the test speech unit segments to a predetermined threshold value; wherein if the combined confidence measure is greater than the predetermined threshold value, the phonetic is accepted (The confidence measure of the syllable verification can effectively detect the fatal reading error. The confidence measure of the segmentation can better detect the rest of (non\-fatal) errors; conclusion.).

9. Consider claim 2, Kuo teaches the method as claimed in claim 1, wherein the segmenting step further comprises:

using a hidden Markov model (HMM) to cut the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script, wherein each test speech unit segment is defined as correspondingly having an initial cutting point (The analysis window is 20 ms with a window shift of 10 ms. The feature vector has 26 dimensions including 12 Mel-scale cepstral coefficients (MFCC), 12 delta-cepstral coefficients, I delta-energy, and 1 delta-delta-energy. The speaker-dependent HMMs are left-to-right models including 100 3-state right context-dependent Initial models, 38 5-state Final models, and a single I-state silence model. At first, we used speaker-independent HMMs as the initial models for training the speaker-dependent HMMs, section 2.1);

performing a fine adjustment on the initial cutting point of the test speech unit segment according to at least one feature factor corresponding to each test speech unit segment and calculating at least one cutting point fine adjustment value corresponding to each test speech unit segment (The boundaries of Mandarin syllables, Initials and Finals are obtained from the state-level boundaries. As mentioned, most of the syllable boundaries are not accurate enough and need to be appropriately adjusted. In the fine adjustment, we calculated zero crossing rates (ZCR) and energies using a 5 ms window with 1 ms shift. In addition, energies of band-pass and high-pass signals were obtained on a speaker-dependent band. These are the features for fine adjustment. According to the phonemic properties of the Initials, the Mandarin syllables are clustered into 7

categories, as shown in Table 2. For each category, multiple rules based on observation and statistics of the mentioned features were designed to further adjust the boundaries. We used so-called multiple expert decision strategy; section 2.2); and

integrating the initial cutting point and the cutting point fine adjustment value of the test speech unit segment to obtain a cutting point of the test speech unit segment (Fine adjustments based on multiple rules are fused with various strategies (voting, weighted-sum, etc.) for each category of syllable types; section 2.2.).

10. Consider claim 3, Kuo teaches the method as claimed in claim 2, wherein the feature factor of the test speech unit segment is a neighboring cutting point of the initial cutting point (In the fine adjustment, we calculated zero crossing rates (ZCR) and energies using a 5 ms window with 1 ms shift. In addition, energies of band-pass and high-pass signals were obtained on a speaker-dependent band. These are the features for fine adjustment; section 2.2).

11. Consider claim 4, Kuo teaches the method as claimed in claim 2, wherein the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment (In the fine adjustment, we calculated zero crossing rates (ZCR) and energies using a 5 ms window with 1 ms shift. In addition, energies of band-pass and high-pass signals were obtained on a speaker-dependent band. These are the features for fine adjustment; section 2.2).

12. Consider claim 5, Kuo teaches the method as claimed in claim 2, wherein the feature factor of the test speech unit segment is an energy value of the test speech unit segment (In the fine adjustment, we calculated zero crossing rates (ZCR) and energies using a 5 ms window with 1 ms shift. In addition, energies of band-pass and high-pass signals were obtained on a speaker-dependent band. These are the features for fine adjustment; section 2.2).

13. Consider claim 6, Kuo teaches the method as claimed in claim 5, wherein the energy value is an energy value of a band pass signal and a high pass signal retrieved from a speaker-dependent band (In the fine adjustment, we calculated zero crossing rates (ZCR) and energies using a 5 ms window with 1 ms shift. In addition, energies of band-pass and high-pass signals were obtained on a speaker-dependent band. These are the features for fine adjustment; section 2.2).

14. Consider claim 7, the method as claimed in claim 2, wherein each cutting point fine adjustment value has a weighted value, and the cutting point of the test speech unit segment is a weighted average of the initial cutting point and the cutting point fine adjustment value (Fine adjustments based on multiple rules are fused with various strategies (voting, weighted-sum, etc.) for each category of syllable types; section 2.2).

15. Consider claim 9, Kuo teaches tThe method as claimed in claim 1, wherein in the phonetic-confidence-measure step, each phonetic confidence measure of the test speech unit segments is:

CMV = min{LLR~, LLRF,0},

Where LLR, = logP(X, J Ho) - logP(X, I H),

LLR~ = logP(X1, I Ho) - logP(X1, [H), X1 is an initial segment of the test speech unit segment, XF is a final segment of the test speech unit segment, H0 is a null hypothesis of the test speech unit segment recorded correctly, H~ is an alternative hypothesis of the test speech unit segment recorded incorrectly, and LLR is a log likelihood ratio (this is covered verbatim in section 3.4).

16. Claims 10-16 and 18 are rejected under 102 as well for the same reasons and 1-7 and 9 as they contain the same limitations.

Claim Rejections - 35 USC § 103

17. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

18. The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

19. Claims 1-3, 5-7, 10-12, and 14-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chou et al. (Corpus-Based Mandarin Speech Synthesis with Contextual Syllabic Units Based on Phonetic Properties) in view of Modi et al. (US Patent 6,125,345).

20. Consider claim 1, Chou teaches an automatic speech segmentation and verification method (Figure 1, automatic segmentation.) comprising:

a retrieving step, for retrieving a recorded speech corpus, the recorded speech corpus corresponding to a known text script, the known text script defining phonetic information with N phonetic units (With an accompanying orthographic transcription, the corpus can be segmented by labeling with the HMMs, page 894, column 1 line 27. Figure 1, step 1 Waveform and transcription are inputted. The transcription of the waveform would inherently contain the N phonetic units of the waveform.);

a segmenting step, for segmenting the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script (With an accompanying orthographic transcription, the corpus can be

segmented by labeling with the HMMs, page 894, column 1 line 27. Figure 1, waveform and transcription are both inputted to SI HHM segmentation step, showing that both would be considered.);

a segment-confidence-measure verifying step, for verifying segment confidence measures of N cutting points of the test speech unit segments to determine if the N cutting points of the test speech unit segments are correct (To evaluate the effects of the whole process, the output after the manual correction is set as the reference. The errors are calculated as the difference between the determined boundaries and the reference boundaries; page 894, column 2, line 40.);

a determining step, for determining acceptance of the phonetic unit by comparing a segment reliability of the test speech unit segments to a predetermined threshold value; wherein if the combined confidence measure is greater than the predetermined threshold value, the phonetic is accepted (To evaluate the effects of the whole process, the output after the manual correction is set as the reference. The errors are calculated as the difference between the determined boundaries and the reference boundaries. The segmentation rate is defined as the percentage of errors within 10ms and 20ms.).

Chou does not specifically teach:

a phonetic-confidence-measure verifying step, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script; and

nor considering the phonetic confidence measures in the determining the acceptance of the phonetic unit.

In the same field of speech verification, Modi teaches:

a phonetic-confidence-measure verifying step, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script (Regardless of the conventional procedure used to train the recognition HMMs 126 and the verification HMMs 134, in operation, the conventional verification subsystem 130 of the automated speech recognition system 100 uses the verification procedure shown in FIG. 5 to determine whether the recognized utterance is accepted or rejected; column 7 line 54. Figure 5 shows keywords and anti-keywords with probabilities that are used to come up with a likelihood ratio. One of ordinary skill in the art could appreciate that the HMMs used for recognition could be limited to the ones used for segmentation, and the same verification principles of Modi would apply.); and

nor considering the phonetic confidence measures in the determining the acceptance of the phonetic unit (the conventional verification subsystem 130 of the automated speech recognition system 100 uses the verification procedure shown in FIG. 5 to determine whether the recognized utterance is accepted or rejected; column 7 line 54.).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the confidence measure of Modi with the segmentation of and verification of Chou in order to allow for assurance that not only is the phonemes segmented in the right place, they are also the correct segments.

21. Consider claim 2, Chou teaches the method as claimed in claim 1, wherein the segmenting step further comprises:

using a hidden Markov model (HMM) to cut the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script, wherein each test speech unit segment is defined as correspondingly having an initial cutting point (For automatic processing, the boundary correction rules are applied instead of the human correction. These prior described rules are based on the knowledge from the observations in human correction procedures. The outputs of SD HMMs are accepted as the initial boundaries; page 894, column 2, line 20.);

performing a fine adjustment on the initial cutting point of the test speech unit segment according to at least one feature factor corresponding to each test speech unit segment and calculating at least one cutting point fine adjustment value corresponding to each test speech unit segment (For automatic processing, the boundary correction rules are applied instead of the human correction. These prior described rules are based on the knowledge from the observations in human correction procedures. The outputs of SD HMMs are accepted as the initial boundaries. The program then searches in a local area for the acoustic features that match the phonetic properties of the units. The features include RMS power, voicing probability and FFT spectrogram derived from ESPS programs. The window sizes are varied from 5ms to 20ms according to the features and phonetic types of units. For example, a 5ms window of RMS power is applied to locate a plosive because there is a short burst of energy when the sound is

released. If the specified acoustic features are not found in that area, the boundary is left no change; page 894, column 2, lines 20-35.); and

integrating the initial cutting point and the cutting point fine adjustment value of the test speech unit segment to obtain a cutting point of the test speech unit segment (the adjusted boundaries are further processed to update the parameters of the SD HMMs. These procedures are re, cursively performed until the average alternation of boundaries is under a threshold; page 894, column 2, line 34.).

22. Consider claim 3, Chou teaches the method as claimed in claim 2, wherein the feature factor of the test speech unit segment is a neighboring cutting point of the initial cutting point (The program then searches in a local area for the acoustic features that match the phonetic properties of the units; page 894, column 2 line 25.).

23. Consider claim 5, Chou teaches the method as claimed in claim 2, wherein the feature factor of the test speech unit segment is an energy value of the test speech unit segment (The features include RMS power, voicing probability and FFT spectrogram derived from ESPS programs; page 894, column 2, line 27.).

24. Consider claim 6, Chou teaches the method as claimed in claim 5, wherein the energy value is an energy value of a band pass signal and a high pass signal retrieved from a speaker-dependent band (The features include RMS power, voicing probability and FFT spectrogram derived from ESPS programs; page 894, column 2, line 27. The

FFT spectrogram is made up of bandpass signals, and a group of FFT coefficients can be considered together to create a highpass signal. These signals would be in fact speaker dependent as speaker dependent HMMs are used.).

25. Consider claim 7, the method as claimed in claim 2, wherein each cutting point fine adjustment value has a weighted value, and the cutting point of the test speech unit segment is a weighted average of the initial cutting point and the cutting point fine adjustment value (These procedures are re, cursively performed until the average alternation of boundaries is under a threshold; page 894, column 2, line 34. This is in effect an average of the initial cutting point and the adjusted value.)

26. Consider claim 10, Chou teaches an automatic speech segmentation and verification system (Figure 1, automatic segmentation.) comprising:

a database for storing a recorded speech corpus, the recorded speech corpus corresponding to a known text script, the known text script defining phonetic information with N phonetic units (With an accompanying orthographic transcription, the corpus can be segmented by labeling with the HMMs, page 894, column 1 line 27. Figure 1, step 1 Waveform and transcription are inputted. The transcription of the waveform would inherently contain the N phonetic units of the waveform. A database would be inherent in order to allow for processing by a computer);

a segmenting unit, for segmenting the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the

known text script (With an accompanying orthographic transcription, the corpus can be segmented by labeling with the HMMs, page 894, column 1 line 27. Figure 1, waveform and transcription are both inputted to SI HHM segmentation step, showing that both would be considered.);

a segment-confidence-measure verifying unit, for verifying segment confidence measures of N cutting points of the test speech unit segments to determine if the N cutting points of the test speech unit segments are correct (To evaluate the effects of the whole process, the output after the manual correction is set as the reference. The errors are calculated as the difference between the determined boundaries and the reference boundaries; page 894, column 2, line 40.);

a determining unit, for determining acceptance of the phonetic unit by comparing a segment reliability of the test speech unit segments to a predetermined threshold value; wherein if the combined confidence measure is greater than the predetermined threshold value, the phonetic is accepted (To evaluate the effects of the whole process, the output after the manual correction is set as the reference. The errors are calculated as the difference between the determined boundaries and the reference boundaries. The segmentation rate is defined as the percentage of errors within 10ms and 20ms.).

Chou does not specifically teach:

a phonetic-confidence-measure verifying unit, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script; and

nor considering the phonetic confidence measures in the determining the acceptance of the phonetic unit.

In the same field of speech verification, Modi teaches:

a phonetic-confidence-measure verifying unit, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script (Regardless of the conventional procedure used to train the recognition HMMs 126 and the verification HMMs 134, in operation, the conventional verification subsystem 130 of the automated speech recognition system 100 uses the verification procedure shown in FIG. 5 to determine whether the recognized utterance is accepted or rejected; column 7 line 54. Figure 5 shows keywords and anti-keywords with probabilities that are used to come up with a likelihood ratio. One of ordinary skill in the art could appreciate that the HMMs used for recognition could be limited to the ones used for segmentation, and the same verification principles of Modi would apply.); and

considering the phonetic confidence measures in the determining the acceptance of the phonetic unit (the conventional verification subsystem 130 of the automated speech recognition system 100 uses the verification procedure shown in FIG. 5 to determine whether the recognized utterance is accepted or rejected; column 7 line 54.).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the confidence measure of Modi with the segmentation of and verification of Chou in order to allow for assurance that not only is the phonemes segmented in the right place, they are also the correct segments.

27. Consider claim 11, Chou teaches the system as claimed in claim 10, wherein the segmenting unit performs the following steps:

using a hidden Markov model (HMM) to cut the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script, wherein each test speech unit segment is defined as correspondingly having an initial cutting point (For automatic processing, the boundary correction rules are applied instead of the human correction. These prior described rules are based on the knowledge from the observations in human correction procedures. The outputs of SD HMMs are accepted as the initial boundaries; page 894, column 2, line 20.);

performing a fine adjustment on the initial cutting point of the test speech unit segment according to at least one feature factor corresponding to each test speech unit segment and calculating at least one cutting point fine adjustment value corresponding to each test speech unit segment (For automatic processing, the boundary correction rules are applied instead of the human correction. These prior described rules are based on the knowledge from the observations in human correction procedures. The outputs of SD HMMs are accepted as the initial boundaries. The program then searches in a local area for the acoustic features that match the phonetic properties of the units. The features include RMS power, voicing probability and FFT spectrogram derived from ESPS programs. The window sizes are varied from 5ms to 20ms according to the features and phonetic types of units. For example, a 5ms window of RMS power is

applied to locate a plosive because there is a short burst of energy when the sound is released. If the specified acoustic features are not found in that area, the boundary is left no change; page 894, column 2, lines 20-35.); and

integrating the initial cutting point and the cutting point fine adjustment value of the test speech unit segment to obtain a cutting point of the test speech unit segment (the adjusted boundaries are further processed to update the parameters of the SD HMMs. These procedures are re, cursively performed until the average alternation of boundaries is under a threshold; page 894, column 2, line 34.).

28. Consider claim 12, Chou teaches the system as claimed in claim 11, wherein the feature factor of the test speech unit segment is a neighboring cutting point of the initial cutting point (The program then searches in a local area for the acoustic features that match the phonetic properties of the units; page 894, column 2 line 25.).

29. Consider claim 14, Chou teaches the method as claimed in claim 11, wherein the feature factor of the test speech unit segment is an energy value of the test speech unit segment (The features include RMS power, voicing probability and FFT spectrogram derived from ESPS programs; page 894, column 2, line 27.).

30. Consider claim 15, Chou teaches the method as claimed in claim 14, wherein the energy value is an energy value of a band pass signal and a high pass signal retrieved from a speaker-dependent band (The features include RMS power, voicing probability

and FFT spectrogram derived from ESPS programs; page 894, column 2, line 27. The FFT spectrogram is made up of bandpass signals, and a group of FFT coefficients can be considered together to create a highpass signal. These signals would be in fact speaker dependent as speaker dependent HMMs are used.).

31. Consider claim 16, the method as claimed in claim 11, wherein each cutting point fine adjustment value has a weighted value, and the cutting point of the test speech unit segment is a weighted average of the initial cutting point and the cutting point fine adjustment value (These procedures are re, cursively performed until the average alternation of boundaries is under a threshold; page 894, column 2, line 34. This is in effect an average of the initial cutting point and the adjusted value.)

32. Claims 4 and 13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chou in view of Modi as applied to claim 2 above, and further in view of Toledano et al (Trying to Mimic Human Segmentation of Speech Using HHM and Fuzzy Logic Post-Correction Rules).

33. Consider claim 4, Chou in view of Modi teaches the method as claimed in claim 2, but does not specifically teach wherein the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment.

In the same field of segmentation verification, Toledano teaches the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech

unit segment (Signal features at a time position are computed based on two windows of fixed width. Among these features is the zero crossing rate; page 4, column 1, paragraph 1.).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the use of zero crossing rate as taught by Toledano with the verification method of Chou and Modi in order to provide another tool for assessing the accuracy of the segmentation.

34. Consider claim 13, Chou in view of Modi teaches the system as claimed in claim 2, but does not specifically teach wherein the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment.

In the same field of segmentation verification, Toledano teaches the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment (Signal features at a time position are computed based on two windows of fixed width. Among these features is the zero crossing rate; page 4, column 1, paragraph 1.).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the use of zero crossing rate as taught by Toledano with the verification method of Chou and Modi in order to provide another tool for assessing the accuracy of the segmentation.

Allowable Subject Matter

35. Claim 8 would be allowable if rewritten to overcome the rejection(s) under 35 U.S.C. 101, set forth in this Office action and to include all of the limitations of the base claim and any intervening claims.

36. Consider claim 8, Kuo nor the combination of Chou and Modi fairly suggests the method as claimed in claim 1, wherein in the segment-confidence-measure step, each segment confidence measure of the test speech unit segment is:

CMS = max(1- h(D)- epsilon g(c(s), f(s)), 0)

where h(D)= K(Epsilon wi |di –di|), D is a vector of multiple expert decisions of the cutting point, di is the cutting point, d = p(D) is a final decision of the cutting point, K(x) is a monotonically increasing function that maps a non-negative variable x into a value between 0 and 1, g(c(s), f(s)) is a cost function value between a cost function ranging from 0 to 1, s is a segment, c(s) is a type category of the segment s and, f(s) are acoustic features of the segment.

37. Claim 17 objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

38. Consider claim 17, Kuo nor the combination of Chou and Modi fairly suggests the system as claimed in claim 10, wherein in the segment-confidence-measure step, each segment confidence measure of the test speech unit segment is:

CMS = max(1- h(D)- epsilon g(c(s), f(s)), 0)

where h(D)= K(Epsilon wi |di –~~d~~i|), D is a vector of multiple expert decisions of the cutting point, di is the cutting point, ~~d~~ = p(D) is a final decision of the cutting point, K(x) is a monotonically increasing function that maps a non-negative variable x into a value between 0 and 1, g(c(s), f(s)) is a cost function value between a cost function ranging from 0 to 1, s is a segment, c(s) is a type category of the segment s and, f(s) are acoustic features of the segment.

Conclusion

39. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure is included on the Notice of References Cited.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Douglas C. Godbold whose telephone number is (571) 270-1451. The examiner can normally be reached on Monday-Thursday 7:00am-4:30pm Friday 7:00am-3:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571) 272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DCG

PATRICK N. EDOUARD
SUPERVISORY PATENT EXAMINER